Naval Research Laboratory
Washington, DC 20375-5320

NRL/FR/5555--13-10,239

Universal Vocoder Using Variable Data Rate Vocoding

David A. Heide
Aaron E. Cohen
Yvette T. Lee
Thomas M. Moran

Transmission Technology Branch
Information Technology Division

June 14, 2013

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
14-06-2013

2. REPORT TYPE
Formal Report

3. DATES COVERED *(From - To)*
October 1, 2010 to February 20, 2013

4. TITLE AND SUBTITLE
Universal Vocoder Using Variable Data Rate Vocoding

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62235N

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
David A. Heide, Aaron E. Cohen, Yvette T. Lee, and Thomas M. Moran

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/FR/5555--13-10,239

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

10. SPONSOR / MONITOR'S ACRONYM(S)

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

To help achieve universal secure communications interoperability in the Department of Defense (DoD), one intermediate goal has been the development of a universal voice encoder (vocoder) that can seamlessly encode speech at a wide range of variable and fixed data rates to suit a wide range of DoD communication equipment. This report describes the most recent advancements in achieving this goal. Four of the most important areas of improvements presented are: (1) Significant improvements were made to the variable data rate (VDR) vocoder that make it much more robust in less than ideal environments. (2) Error control coding is now extended to all VDR modes. (3) Fixed rate vocoding modes based directly on the VDR encoding method were designed so that transcoding between these options can be done directly and with very little degradation in voice quality. (4) Heavily error protected, fixed rate modes were designed. These modes can be used as fail-safe modes to ensure communicability when channels deteriorate to previously unusable levels.

15. SUBJECT TERMS
Variable data rate vocoder   MELPe vocoder   Transcoding
Speech modeling   Error control coding

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			David A. Heide
Unclassified	Unclassified	Unclassified	Unlimited	39	**19b. TELEPHONE NUMBER** *(include area code)* 202-404-7107

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

CONTENTS

1 INTRODUCTION 1

1.1 Part One, Description of the VDR Algorithm and Significant Improvements Since 2007 (Section 2) 1
1.2 Part Two, Extension of Error Control Coding in VDR Modes to Cover Many More Voice Applications (Section 3) 1
1.3 Part Three, Addition of Fixed Rate Options that can be Transcoded to/from VDR Modes to Achieve Universal Interoperability (Sections 4, 5, 6) 2

2 VDR VOCODING ALGORITHM 3

2.1 Background of VDR 3
2.2 Generation of the Speech Prediction Residual 4
2.3 Encoding the Prediction Residual Spectrum 6
2.4 Description of Quantization Tables for VDR 7

3 PROVIDING VARYING BIT ERROR PROTECTION FOR VDR MODES 13

3.1 Description of All VDR Modes with Varying Levels of ECC 13
3.2 Quantization Tables for All the VDR Modes 14
3.3 Mode Switching 19

4 EXTENDING VDR TO 16000 BPS FIXED RATE OPTIONS (WITH AND WITHOUT ECC) 22

4.1 Description of 16000 bps Fixed Rate Modes 22

5 TRANSCODING 16000 BPS FIXED RATE MODES TO/FROM VARIABLE RATE MODES 24

5.1 VDR Modes and Fixed Rate Modes to be Transcoded 24
5.2 Conversion Between Different Precision of Spectral Constellations 25
5.3 Testing Sample 25
5.4 Transcoding from Variable Data Rate Modes to Fixed Rate Modes 26
5.5 Transcoding from Fixed Rate Modes to Variable Data Rate Modes 29

6 DESIGNING FIXED RATE 8000, 12000, 600, AND 1200 MELPE MODES WITH ECC INTO BIT ERROR TOLERANT MODES 31

6.1 8000 and 12000 bps Fixed Rate Modes Based on 2400 bps MELPe Option 31
6.2 2400 bps Fixed Rate Modes Based on 1200 and 600 bps MELPe Vocoding Options 32

7 CONCLUSIONS 33

ACKNOWLEDGMENTS 33

REFERENCES 34

UNIVERSAL VOCODER USING VARIABLE DATA RATE VOCODING

1 INTRODUCTION

In 2007, the Voice Systems Section of the Naval Research Laboratory (NRL) published a report titled "Variable Data Rate Voice Encoder for Narrowband and Wideband Speech" [1]. In this report, we described a voice coder (vocoder) that, based on both speech content and external network constraints, encoded speech at dynamically varying data rates. The initial NRL variable data rate (VDR) vocoder concept was documented in 2001 by George Kang [2].

A single voice processing principle is used to generate the various data rates in NRL's vocoder. This feature of the vocoder algorithm allows voice encoded at different rates to be interoperable. So, for example, voice can be encoded at a high rate when the channel bandwidth is available, but the rate can be reduced in mid-transmission if the channel bandwidth becomes restricted. The receiving voice terminal will always be able to decode the voice, regardless of the change in rate. This can happen without external or prior signaling and as often as every 22.5 milliseconds (ms). This feature allows voice quality to be constantly balanced with the available channel bandwidth. It also allows for voice over high bandwidth channels to be directly interoperable with voice over narrow bandwidth channels and vice versa.

While this 2007 work significantly advanced of the state of the art, we have added many capabilities and improvements to the VDR vocoder since then as part of NRL's work toward developing a universal vocoder for the Department of Defense (DoD). The present report documents these advancements in three main parts, as outlined below.

1.1 Part One, Description of the VDR Algorithm and Significant Improvements Since 2007 (Section 2)

To detail the significant improvements in the VDR algorithm since 2007, Section 2 reviews the algorithm and lessons learned from recent testing. One lesson involved improving the speech analysis with a much more robust way of determining each speech frame's optimum level of encoding precision. This resulted in a significant increase in voice quality when speech is recorded under less than ideal conditions. A second lesson improved upon the speech synthesis technique of the receiver for improving the quality of the generated speech without any corresponding increase in data rate.

1.2 Part Two, Extension of Error Control Coding in VDR Modes to Cover Many More Voice Applications (Section 3)

While VDR was originally developed for varying the encoding rate based on speech content and network congestion, we soon realized it could be extended to many more voice applications by including a selection of error correcting codes within each mode. Section 3 describes including error control coding (ECC) within VDR so that modes can be tailored to specific channel environments or specific acoustic environments and to make it possible to automatically switch between modes to optimize performance. For example, difficult channel environments can use modes with more ECC while difficult acoustic

Manuscript approved May 20, 2013.

environments can use modes that are less susceptible to acoustic noise. The goal is to provide enough possible modes so that we can continuously select a mode that gives good overall voice quality given the current conditions.

1.3 Part Three, Addition of Fixed Rate Options that can be Transcoded to/from VDR Modes to Achieve Universal Interoperability (Sections 4, 5, 6)

Fixed rate options were added so that interoperability can be achieved across a wide range of communication devices through transcoding. For this report, transcoding is the process of converting encoded voice with one vocoder to encoded voice with another vocoder. Sections 4, 5, and 6 cover this topic.

1.3.1 Extend VDR to Fixed Rate Options with and without ECC (Section 4)

Many DoD platforms require fixed rate modes. Section 4 describes these modes that are designed for fixed rate radio applications such as HF/UHF. NRL developed two fixed rate variants specifically designed for 16000 bps channels, one with ECC and one without. In addition, like all VDR modes, these fixed rate modes were designed to be directly compatible with the DoD and North Atlantic Treaty Organization (NATO) narrowband vocoder, Mixed Excitation Linear Prediction enhanced (MELPe) [3], ensuring interoperability over the most disadvantaged channels.

1.3.2 Present Techniques and Results of Transcoding between Fixed Rate and Variable Rate Modes (Section 5)

Section 5 presents the techniques that make it possible to directly cross multiple different links to reach the end user without harming voice quality. To achieve direct interoperability, NRL designed these fixed rate modes with the same speech analysis as VDR, so as a result, these modes can be directly transcoded to/from VDR modes. Because all modes are derived from the same voice processing principle, the conversion process eliminates the complete decoding of all voice parameters and then re-encoding with the new vocoding algorithm, which can significantly degrade voice quality. To convert the variable to the fixed rate vocoder and vice versa, only one voice parameter set needs to be transcoded. This parameter is the "prediction residual," discussed in Section 2. The prediction residual is particularly important because the "variable" feature in VDR comes directly from changing the precision in encoding it.

1.3.3 Design Fixed Rate Modes Based on MELPe that are not Dependent on VDR (Section 6)

Section 6 describes four fixed rate vocoders that are dependent on various MELPe modes. These modes do not depend on the VDR speech analysis but are presented as part of a suite of modes possible to make a truly universal coder. Included in these options are two 2400 bps modes, an 8000 bps mode, and a 12000 bps mode. These modes include significant levels of ECC to make robust vocoders in severe channel environments.

2 VDR VOCODING ALGORITHM

2.1 Background of VDR

2.1.1 Benefits of Using a Single Voice Processing Principle in VDR

In the past, most communications equipment was designed and procured individually without much regard to interoperability. The design of each communication system was often limited to the individual communication link. RF link distances and quality vary. The techniques for reliably transmitting secure voice have also varied and have been specific to the individual link and that link's data rate. While this approach ensured that each link was designed for optimum performance, absolutely no interoperability across different links was possible without completely decoding the speech, synthesizing it, reanalyzing it, and finally, re-encoding it.

To address the need for interoperability, NRL designed VDR to operate over a wide range of rates and to easily change rates on the fly. This way there is no need to implement several different vocoders each running at a different, incompatible rate. VDR uses a single voice processing principle to operate over a wide variety of data rates, all of them interoperable, and with the instantaneous rate constantly changing to the optimal rate, based on a variety of inputs.

In addition, NRL designed VDR to be based on the NATO 2.4 kbps standard vocoder, MELPe. VDR improves upon MELPe by encoding the excitation signal with finer and finer precision of the speech prediction residual. Basing VDR on MELPe was a very important decision for several reasons:

- MELPe bitstream could be embedded in the VDR bitstream so that all modes would be interoperable;
- MELPe was tested in a wide range of acoustic noise environments that are present in the military;
- MELPe has a noise canceling (NC) preprocessor built in (very beneficial to improving performance in difficult acoustic noise environments);
- MELPe has 600 and 1200 bps options (discussed in Section 6 for designing bit error tolerant modes).

2.1.2 VDR Uses a Two-Dimensional Coding to Vary the Data Rate

VDR encodes speech using two main criteria to decide the amount of precision to use:

1. Speech content. Because vowels are more complex than consonants or gaps, VDR uses more precision (data rate) to encode vowels. For each 22.5 ms frame, VDR can choose one of six levels of precision based on speech content.

2. Network capacity. VDR also has the option of increasing or decreasing the overall data rate based on channel capacity. VDR uses five different overall modes (each with six submodes) in addition to the 2.4 kbps MELPe standard.

By combining speech content options (six) and network capacity options (five), the instantaneous speech content can be encoded with 31 total options (including fixed rate MELPe). This two-dimensional coding is shown in Fig. 1.

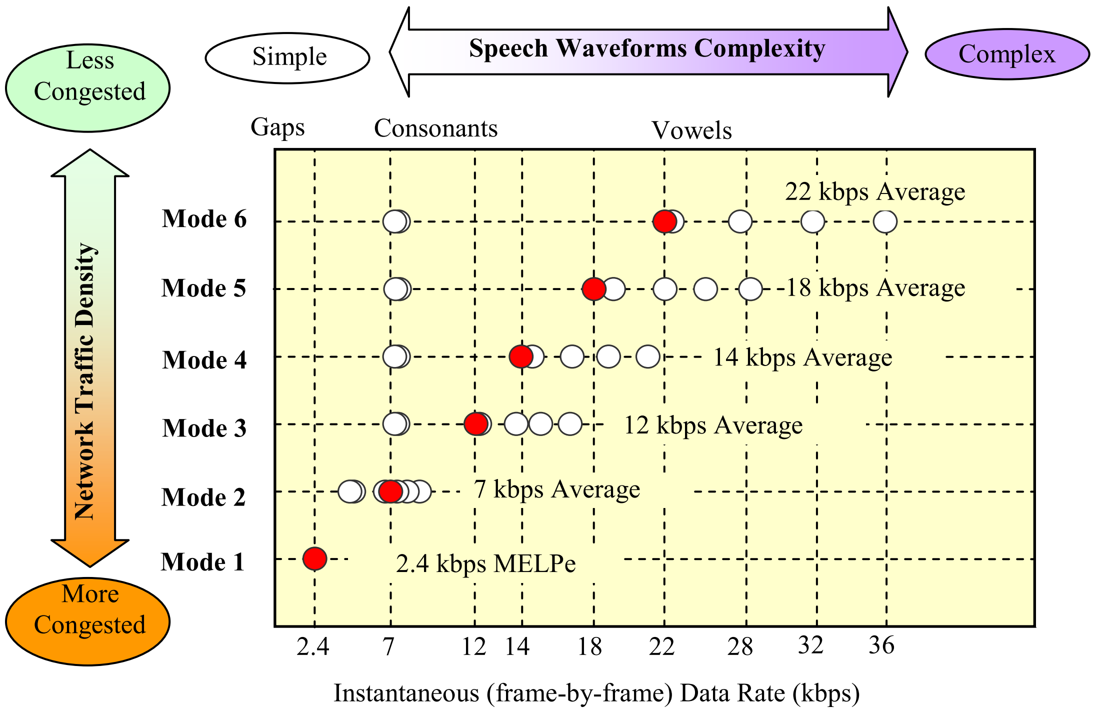


Fig. 1 — Two-dimensional optimization of data rates based on network traffic conditions and the complexity of the speech waveform. Modes 2 through 6 each have six possible submodes, giving 31 total modes including the MELPe standard mode 1. The red circle is the average data rate of each mode.

2.2 Generation of the Speech Prediction Residual

The heart of the "variable" in the VDR coding algorithm derives from the variable precision in the prediction residual encoding process. This process is described more completely in Ref. 1; we summarize the residual encoding process here. A block diagram of the VDR encoding/decoding process is shown in Fig. 2.

The VDR analyzer is divided into three main stages: a two-stage spectral whitening (flattening) process followed by the residual encoder. The first stage attenuates speech resonant frequencies and the second stage attenuates pitch harmonics. The third stage is the residual encoder itself. The first two stages are similar to most linear predictive coding (LPC)-based encoders in which the system decomposes the speech waveform into slowly time-varying components and fast time-varying components. The slowly time-varying components include LPC filter coefficients, pitch value, and speech loudness. They are updated only once per frame (22.5 ms). The fast time-varying components are the prediction residual samples. They are updated sample by sample, 8000 times per second (or every 125 µs). Note that even if the slowly time-varying components are quantized, as long as the prediction residual samples are computed from the quantized slowly time-varying components, the output speech quality is dependent solely on the resolution of the prediction residual.

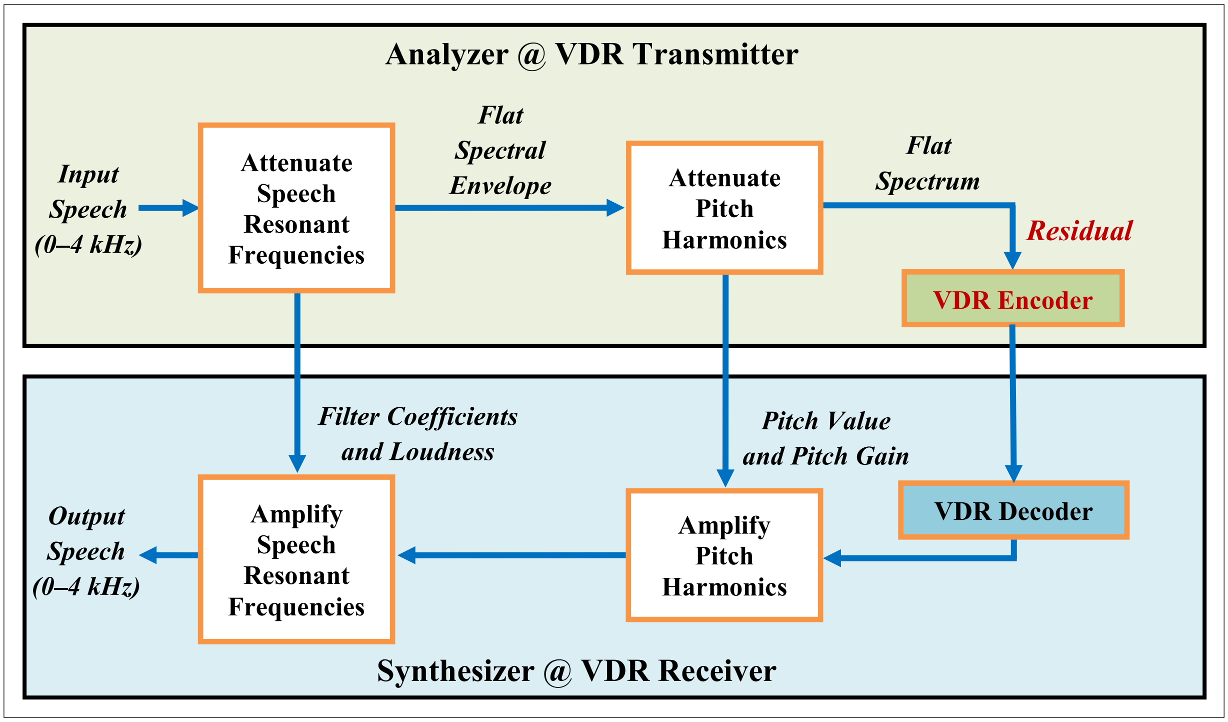


Fig. 2 — Block diagram of VDR based on the linear predictive coding (LPC) analysis/synthesis system. The output speech quality is dependent solely on the resolution (the number of bits used to encode) of the residual, highlighted in red.

Figure 2 shows that the output of the second stage is the prediction residual (highlighted in red) and thus, the data rate of the VDR system and the output speech quality can be controlled by the number of bits used to encode the prediction residual. The output speech quality improves as the resolution of the error signal (the prediction residual) becomes finer (i.e., encoded at a higher data rate). At the finest level of resolution, the system generates an output signal that equals the input. In other words, this one system component is responsible for encoding speech at widely varying rates with correspondingly varying levels of speech quality.

One of the advantages of the VDR system is its flexibility. Not only can it constantly change the data rate based on the complexity of the speech signal, it also is flexible based on external network requirements. So if, for example, an aggregate channel has capacity for a fixed *total* data rate, the encoding rate of the users can be adjusted based on how many users are communicating at any given time.

To ensure compatibility with the MELPe 2.4 kbps standard vocoder, the exact 54-bit MELPe bitstream is used as the base kernel of the VDR bitstream. We are able to use common parameters from MELPe to save bits in the VDR portion of the bitstream because MELPe and VDR are both based on linear predictive coding. The common parameters used are the LPC parameters (in the form of line spectral pairs) and the pitch.

2.3 Encoding the Prediction Residual Spectrum

The VDR residual encoder operates in the frequency domain. To derive the spectrum of the residual, each 180 speech sample frame is overlapped with 12 samples of the previous frame. The resulting 192 samples are windowed and then are transformed using the Winograd transform. This process generates 96 complex (real and imaginary) spectral coefficients that represent the entire 0 to 4 kHz audio spectrum. The DC component and the first spectral component (at $f = 41.67$ Hz) are not transmitted because they do not result in audible sounds. *The data rate of the VDR residual encoder is completely dependent on and varied by how many of the remaining 94 coefficients are encoded and the precision of each coefficient.*

Figure 3 provides an example of the entire spectrum of the prediction residual. The graph represents the amplitude of the 4 kHz speech residual. The complete VDR system uses 94 of the spectral coefficients covering the 100 to 4000 Hz bandwidth.

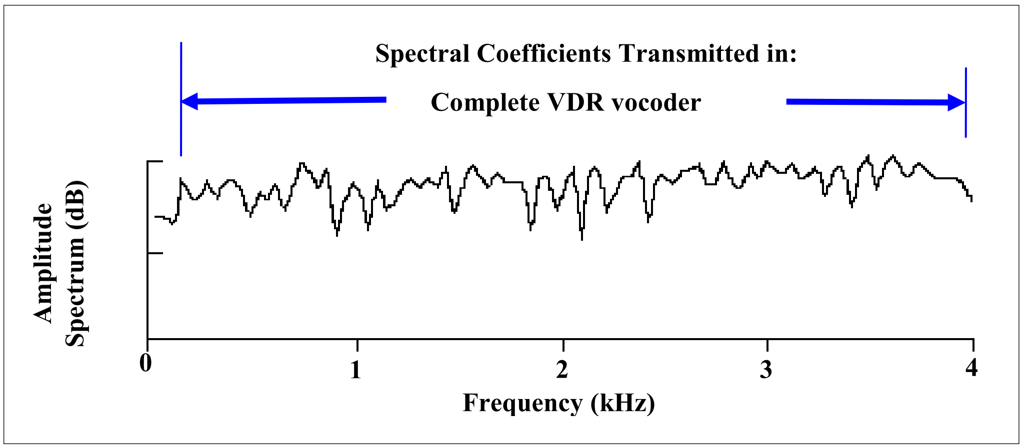
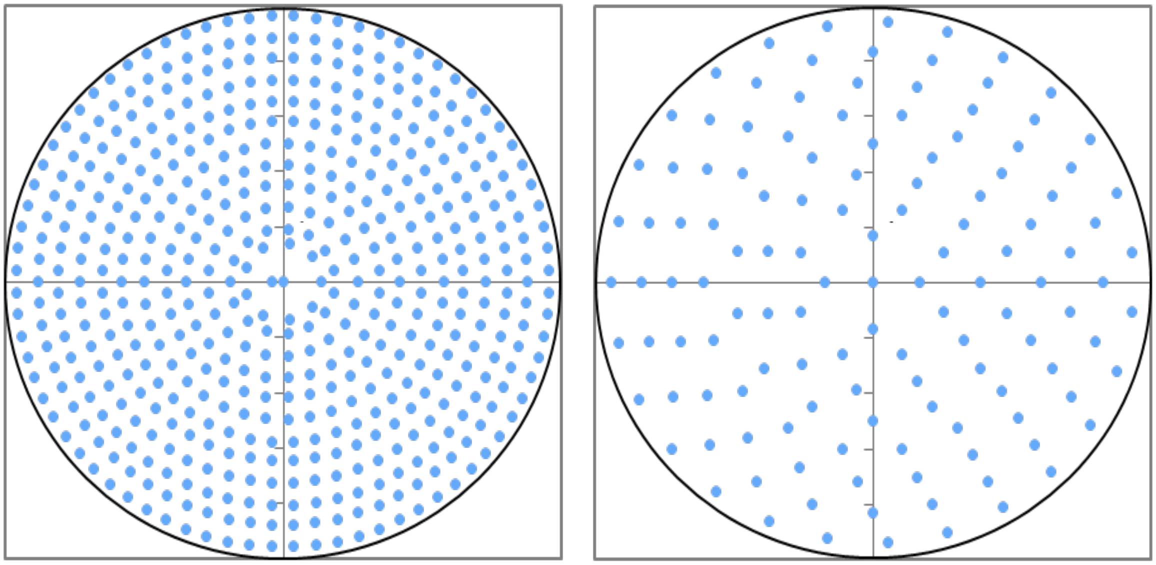


Fig. 3 — Example of the 4 kHz, 96-point residual spectrum. The complete VDR system encodes the 100 to 4000 Hz bandwidth.

To encode the 94 coefficients, the real and imaginary coefficients of *each* spectral coefficient are mapped into the unit circle. The data rate is then determined by how many bits are used to “cover” the entire unit circle. For example, a 9-bit table forms a constellation of 512 different spectral codes. A 7-bit table forms a constellation of 128 different spectral codes. Figure 4 illustrates these two spectral encoding constellations as examples. The complete VDR encoder uses five different coding tables (9-bit, 8-bit, 7-bit, 6-bit, and 3-bit tables) to vary the data rate. With the LPC analysis/synthesis method, if the entire spectrum is left unquantized, there is no degradation. The degradation in voice quality comes from the difference (error) between the unquantized residual coefficients and the quantized values that are represented in the constellation. The greater the number of spectral codes in the constellation, the smaller the quantization error and the less the degradation. One of the most important design features of the algorithm is determining when to use more bits (when the error can be heard) and when to use fewer bits (when the error is not audible.)



(a) 9-bit spectral constellation

(b) 7-bit spectral constellation

Fig. 4 — Examples of two spectral coding constellations, 9-bit and 7-bit. Here, the spectral coefficients are quantized jointly by amplitude and phase and are represented on a unit circle. The 9-bit constellation in (a) uses 512 points to cover the unit circle. Because there are a relatively large number of points, the difference (error) between the unquantized spectral coefficient and the quantized value is relatively small. The 7-bit spectral constellation in (b) uses 128 points to cover the unit circle, giving a higher error in quantizing the spectral coefficient in comparison with the 9-bit constellation in (a).

2.4 Description of Quantization Tables for VDR

To reduce the data rates from the highest levels, VDR uses three different techniques.

2.4.1 Conserving Data Based on Speech Complexity

One way to conserve data is by using the variable nature of the speech signal itself. It has long been known that vowels (voiced speech) need much more resolution than consonants (unvoiced speech) or silence do. Figure 5 shows the waveform of a speaker uttering the word “strong.” Notice how complex the waveform is during the “o” vowel, but the consonant “s” at the beginning is little more than random noise. While fixed rate vocoders would encode all these frames with the same precision, VDR analyzes each 22.5 ms frame and decides on the appropriate precision. Past versions of VDR used a spectral complexity index based on the complexity of the prediction residual to determine the appropriate precision for each frame. *The newest version of VDR has completely updated this parameter to a voicing based spectral complexity index.* The reason for changing to voicing based spectral complexity index is that the previous spectral complexity index was sometimes affected by nonrelevant input parameters and encoded at too low a precision given the input speech complexity. Voicing is a measure of correlation in a speech frame. Complex waveforms like vowels are considered voiced and consonants are considered unvoiced. Old vocoders made only one overall determination of voicing for each frame, but MELPe calculates the voicing decision in five separate frequency bands (0–500, 500–1000, 1000–2000, 2000–3000, and 3000–4000 Hz). *Based on extensive testing, we were able to significantly improve performance by changing to this voicing based spectral complexity index.* VDR uses these five voiced/unvoiced decisions to decide how many bits to encode the frame. By summing up the number of frequency bands

that are voiced, MELPe gives us six different degrees of voicing in the speech signal (0 frequency bands voiced up to all 5 frequency bands voiced). This is shown by reading Table 2 vertically up and down. (The VDR encoding tables are in Section 2.4.4.) Note how the top level (five frequency bands voiced) is encoded at a maximum of 808 bits per frame, while frames with 0 or 1 voiced bands are encoded only at 172 bits per frame. This variability in bit precision based on speech complexity is the central way that VDR is made variable.

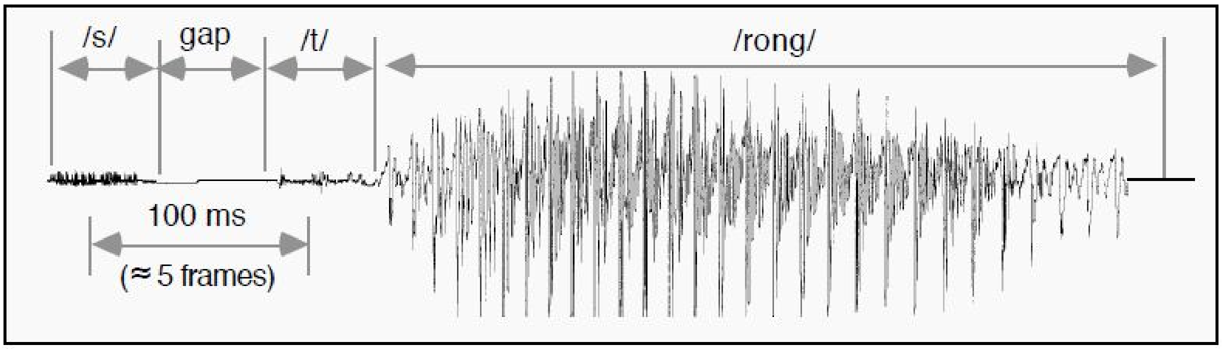


Fig. 5 — Waveform of the word "strong." Notice how the consonant "s" is essentially random noise, while the vowel "o" is a very complex waveform. By analyzing the waveform 44.44 times per second, the VDR algorithm determines the appropriate level of precision to encode the spectral coefficients.

2.4.2 Conserving Data by Using Less Resolution for Higher Frequency Components

A second way to conserve data is by taking advantage of the ear's decreased sensitivity to higher frequencies. Based on earlier studies, it is known that the human ear gradually loses frequency resolution capability for higher frequencies [2]. Therefore, we allow coarser quantization for higher frequency spectral components. Table 2 shows how this fact is utilized by noting the coefficient precision as the table reads horizontally left to right (increasing frequency). Note that the components in the 100 to 1500 Hz band use one more bit resolution than in the 1500 to 2000 Hz band, and two more bits resolution than the 2000 to 4000 Hz band. Encoding the higher frequencies of the speech content less accurately (using fewer bits) than the lower frequencies results in a lower overall data rate than if VDR encoded all coefficients at the higher precision.

2.4.3 Conserving Data by Using Subsets of the Complete VDR Table

A third way to conserve data is by using lower data rate VDR modes that use only subsets of the complete VDR table. In other words, some of the upper frequency band coefficients are completely discarded and replaced either by spectral replication or by inserting the signal derived from the original MELPe upper band. This allows for overall lower rate modes that may be necessary based on channel capacity conditions.

One of the ways VDR is able to lower the data rate for some of the speech modes is to use spectral replication in the spectral coefficients defining the residual excitation signal. That is, after the speech signal has been filtered by the inverse LPC filter and the inverse pitch filter, the resulting signal is analyzed with a 192-point fast Fourier transform (FFT). In the highest data rate mode, all coefficients are

quantized and encoded. For lower rate modes, not all spectral coefficients can be sent. Spectral replication (from lower frequency coefficients to higher rate coefficients) is used at the receiver to closely replicate the excitation signal. Figure 6 shows the frequency range of the spectral coefficients sent in each mode. Table 1 shows the resulting average data rates of these modes. Note that mode 1 is exactly the standardized MELPe algorithm selected for use in the DoD and NATO at 2400 bps.

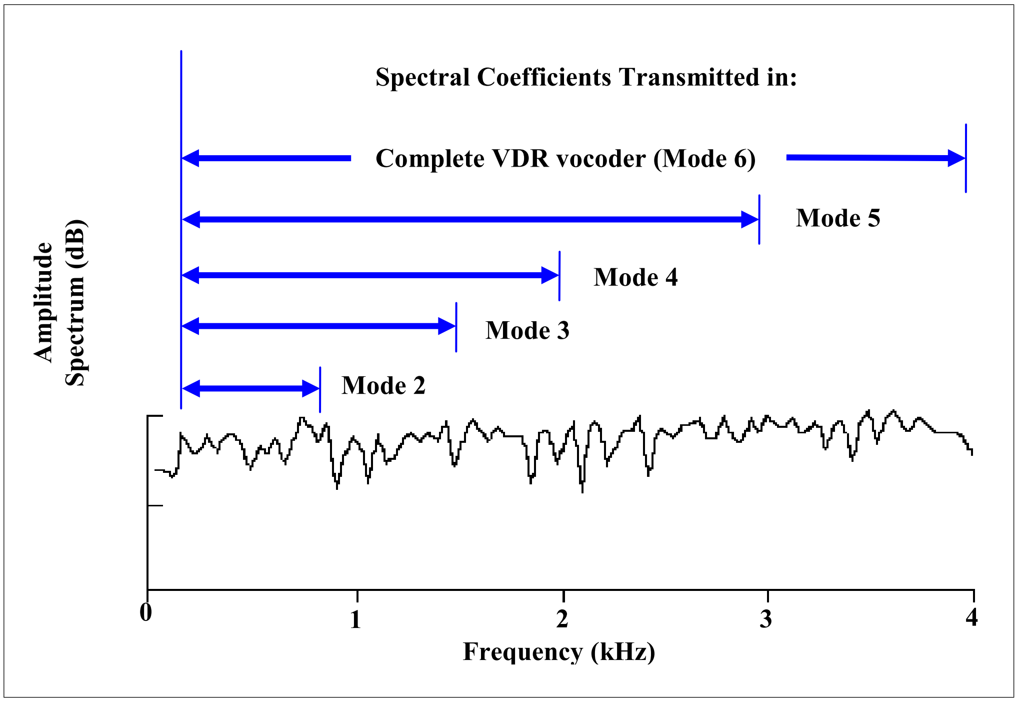


Fig. 6 — Example of the 4 kHz, 96-point residual spectrum and the portion used for each given operating mode, as indicated

Table 1 — VDR Operating Modes

Mode #	**Description**	**Average Mode Data Rates**
Mode 1	MELPe Standard	2.4 kbps Fixed
Mode 2	Hybrid of VDR with MELPe signal above 0.7 kHz	7 kbps
Mode 3	Hybrid of VDR with MELPe signal above 1.5 kHz	12 kbps
Mode 4	VDR with spectral replication above 2 kHz	14 kbps
Mode 5	VDR with spectral replication above 3 kHz	18 kbps
Mode 6	VDR with no spectral replication	22 kbps

The ability of spectral replication of the FFT coefficients to replicate the residual excitation signal diminishes as the frequency band increases. For this reason it is not done for mode 2. In that mode there are only enough vocoder bits available in transmission to cover the first 700 Hz of the residual excitation signal. Since, at the receiver, spectral replication would perform poorly for the remaining 700 to 4000 Hz signal, that relatively broad portion of the spectrum is covered by using the 2400 bps MELPe upperband residual excitation that is transmitted as part of the mode 1 kernel.

Recently, through extensive formal voice intelligibility and acceptability testing, it was found that VDR performance can be improved by also using the MELPe upperband residual excitation signal in mode 3 in addition to mode 2. In past versions of VDR mode 3, the coefficients for the 0 to 1500 Hz band were transmitted and spectral replication was used to replicate the 1500 to 4000 Hz band at the receiver. It was found that a small but consistent improvement in voice quality can be achieved by using the MELPe upperband (1500 to 4000 Hz) instead of spectral replication for this mode.

In modes 4 and 5, the spectral coefficients cover the first 2000 Hz or 3000 Hz, respectively, so the lowerband portion of the spectrum only needs to be replicated once to cover the remaining upperband. So spectral replication provides better voice quality than using the MELPe upperband for the excitation residual in these two modes. In mode 6, the entire frequency band for the excitation residual is transmitted, so neither spectral replication nor the MELPe upperband residual is needed.

2.4.4 VDR Quantization Tables

Table 2 gives the bit allocation for encoding the complete VDR spectrum (mode 6) where all 94 spectral coefficients are encoded and transmitted. (Recall that the first two coefficients are not sent because very low frequencies near DC are not important for speech quality.) Tables 3 through 6 show modes 5, 4, 3, and 2, which use decreasing subsets of the complete coding table found in Table 2. Note that while speech frames with 0 or 1 voice frequency bands both encode the frame with an identical number of bits, they are encoded as separate modes to ensure that future versions of the algorithm can accommodate different precision levels for these two cases.

Table 2 — Mode 6 VDR Quantization Table (Complete Full Rate VDR)

Number of Voiced Frequency Bands		Frequency Band in kHz (# of bits multiplied by # of Spectral Components)				Total # of Bits (note 2)	Instantaneous Data Rate (kbps)
		0.1–1.5 kHz (34)	1.5–2.0 kHz (12)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×34=306	8×12=96	7×24=168	7×24=168	808	35.9
	4	8×34=272	7×12=84	6×24=144	6×24=144	714	31.7
	3	7×34=238	6×12=72	5×24=120	5×24=120	620	27.6
	2	6×34=204	5×12=60	4×24=96	4×24=96	526	23.4
Fully Unvoiced	1	3×34=102	0 (note 1)	0 (note 1)	0 (note 1)	172	7.6
	0	3×34=102	0 (note 1)	0 (note 1)	0 (note 1)	172	7.6

Note 1: The 0 bit means random noise having a unit variance is used for excitation.
Note 2: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 3 — Mode 5 VDR Quantization Table

Number of Voiced Frequency Bands		Frequency Band in kHz (# of bits multiplied by # of Spectral Components)				Total # of Bits (note 3)	Instantaneous Data Rate (kbps)
		0.1–1.5 kHz (34)	1.5–2.0 kHz (12)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×34=306	8×12=96	7×24=168	Not trans-mitted (note 2)	640	28.4
	4	8×34=272	7×12=84	6×24=144		570	25.3
	3	7×34=238	6×12=72	5×24=120		500	22.2
	2	6×34=204	5×12=60	4×24=96		430	19.1
Fully Unvoiced	1	3×34=102	0 (note 1)	0 (note 1)		172	7.6
	0	3×34=102	0 (note 1)	0 (note 1)		172	7.6

Note 1: The 0 bit means random noise having a unit variance is used for excitation.
Note 2: The untransmitted spectral components are replicated by the transmitted spectra in the lower bands.
Note 3: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 4 — Mode 4 VDR Quantization Table

Number of Voiced Frequency Bands		Frequency Band in kHz (# of bits multiplied by # of Spectral Components)				Total # of Bits (note 3)	Instantaneous Data Rate (kbps)
		0.1–1.5 kHz (34)	1.5–2.0 kHz (12)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×34=306	8×12=96	Not transmitted (note 2)		472	21.0
	4	8×34=272	7×12=84			426	18.9
	3	7×34=238	6×12=72			380	16.9
	2	6×34=204	5×12=60			334	14.8
Fully Unvoiced	1	3×34=102	0 (note 1)			172	7.6
	0	3×34=102	0 (note 1)			172	7.6

Note 1: The 0 bit means random noise having a unit variance is used for excitation.
Note 2: The untransmitted spectral components are replicated by the transmitted spectra in the lower bands.
Note 3: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 5 — Mode 3 VDR Quantization Table

Number of Voiced Frequency Bands		Frequency Band in kHz (# of bits multiplied by # of Spectral Components)				Total # of Bits (note 2)	Instantaneous Data Rate (kbps)
		0.1–1.5 kHz (34)	1.5–2.0 kHz (12)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×34=306	Not transmitted MELPe used above 1.5 kHz (note 1)			376	16.7
	4	8×34=272				342	15.2
	3	7×34=238				308	13.7
	2	6×34=204				274	12.2
Fully Unvoiced	1	3×34=102				172	7.6
	0	3×34=102				172	7.6

Note 1: The 1.5–4.0 kHz is derived not from spectral replication but from that region of the 2.4 kbps MELPe signal.
Note 2: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.

Table 6 — Mode 2 VDR Quantization Table

Number of Voiced Frequency Bands		Frequency Band in kHz (# of bits multiplied by # of Spectral Components)				Total # of Bits (note 2)	Instantaneous Data Rate (kbps)
		0.1–0.7 kHz (14)	0.7–2.0 kHz (32)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×14=126	Not transmitted MELPe used above 0.7 kHz (note 1)			196	8.7
	4	8×14=112				182	8.1
	3	7×14=98				168	7.5
	2	6×14=84				154	6.8
Fully Unvoiced	1	3×14=42				112	5.0
	0	3×14=42				112	5.0

Note 1: The 0.7–4.0 kHz is derived not from spectral replication but from that region of the 2.4 kbps MELPe signal.
Note 2: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.

3 PROVIDING VARYING BIT ERROR PROTECTION FOR VDR MODES

One of the most important additions to the VDR algorithm is in providing different levels of error control coding to all of the various VDR modes. Previously, while the VDR algorithm was adaptable to different rates based on speech composition and network congestion, nothing was provided to help adapt to various channel environments. A universal vocoder should take into account the channel characteristics that might be encountered and be able to adapt to them on the fly. In this design, we are using block (frame)-based ECC. These error-protected modes have several benefits:

- *Many modes can be tailored to various channel characteristics*. Each individual VDR mode has four different levels of error protection, so the algorithm can be variable with respect to channel quality.

- *Switching between ECC modes within a common voice mode can be immediate*. The memoryless nature of block (frame) ECC allows the modes to be varied 44.44 times per second, just like the vocoding adaption rate, so we can vary the vocoder and ECC allocation if, for example, the channel quality suddenly degrades or more communicators begin using the available overall channel bandwidth.

- *Block (frame)-based ECC allows for significant flexibility in broadcasting to many different receivers at one time*. With this frame-based approach, intermediate nodes can simply correct errors in each frame, strip off the ECC bits, and forward the vocoder bitstream to many different individual receivers with the same or new ECC encoding added. Each link can choose this mode based on its individual channel capacity/channel quality. This is how communicators can reach across networks to communicate with a wide variety of channel capacity, error characteristics, number of links, etc., securely. It is also how a small amount of errors on each individual link do not magnify over the whole transmission path because they can be corrected before sending onto the next link in the chain. The memoryless feature of block encoding also ensures that burst errors or frame erasures are not propagated to later frames.

- *ECC is added after encryption, so intermediate network nodes can correct the bitstream securely before transmission over each individual link*. By using ECC on the encrypted bitstream, the bitstream does not need to be de-encrypted at each link to correct bit errors, so that end-to-end security is still possible across networks.

3.1 Description of All VDR Modes with Varying Levels of ECC

As introduced above, each VDR mode will have four error control options (submodes): no error protection, low protection, medium protection, and high protection. Block-based Bose Chaudhuri Hocquenghem (BCH) codes are used for the ECC. These BCH codes take in a segment of information bits, compute the parity bits, and then append them to make a codeword. Each codeword is independent from past and future codewords. BCH codes are defined by (n,k,t):

- n = number of total bits = n = 2m − 1 for m = 3,4,5 ... and shortened versions thereof
- k = number of information bits
- n − k = number of parity (ECC) bits
- t = maximum number of possible bits corrected in a codeword

Shortened versions of these codes are achieved by stripping unnecessary information bits.

In the first error control option (no error protection), none of the bits in the bitstream are protected. In the three submodes with ECC (low, medium, high protection), the first 54 MELPe bits and mode index (8 bits) are always protected with four blocks of BCH (n=31,k=16,t=3) encoding. The spectral coefficients of VDR are protected with varying levels of ECC:

- No error protection
- Low error protection: BCH (n=63,k=51,t=2) on spectral coefficients, BCH (n=31,k=16,t=3) on MELPe bits and mode index
- Medium error protection: BCH (n=63,k=39,t=4) on spectral coefficients, BCH (n=31,k=16,t=3) on MELPe bits and mode index
- High error protection: BCH (n=63,k=30,t=6) on spectral coefficients, BCH (n=31,k=16,t=3) on MELPe bits and mode index

So, for example, in the low error protection submode, 2 bit-errors can be corrected in each 63-bit BCH block protecting the spectral coefficients, and 3 bit-errors can be corrected in each 31-bit BCH block protecting the MELPe bits and mode index. Note that if there are too many errors in the VDR spectral coefficients, they do not need to be processed into speech at the receiver. Since the MELPe bits are protected more strongly, sometimes better voice quality is achieved by discarding the corrupted VDR spectral coefficients and processing only the MELPe bits.

Even further, in really bad channel environments, there is the option to transition to VDR modes with even higher levels of error control. Fixed rate modes such as the 16000 bps with ECC (described in Section 4) or the 8000 bps or 12000 bps with ECC fixed rate modes (described in Section 6) provide even higher coding gain, through the use of higher reliability ECC, to ensure communicability of the 54 MELPe bits under severe channel conditions.

3.2 Quantization Tables for All the VDR Modes

Tables 7 through 11 are quantization tables that correspond to the five VDR modes. These tables include the ECC submodes.

Each table shows the six levels of spectral encoding possible for that table’s VDR mode, the four ECC options possible for each level of spectral encoding, and how these two combinations multiply out to 24 possible instantaneous bit rates for a frame of speech in that VDR mode. In total, these five tables describe 120 different encoding options for a 22.5 ms frame of speech. Note again that while speech frames with 0 or 1 voice frequency bands both encode the frame with an identical number of bits, they are encoded as separate modes to ensure that future versions of the algorithm can accommodate different precision levels for these two cases.

The next section describes how to best switch between the VDR modes to give optimal performance given the channel conditions.

Table 7 — Mode 6 Quantization Table for Narrowband VDR with ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components)				Total # of VDR Bits (note 2)	Number of VDR ECC Bits for Four ECC Strength Options (t=0,2,4,6)	Total # of Bits (note 3)	Instantaneous Total Bit Rate (kbps)
		Band 1 0.1–1.5 kHz (34)	Band 2 1.5–2.0 kHz (12)	Band 3 2.0–3.0 kHz (24)	Band 4 3.0–4.0 kHz (24)				
Complex Waveform	5	9×34= 306	8×12= 96	7×24= 168	7×24= 168	746	0 (no ECC)	808	35.9
							15×12 = 180	1048	46.6
							20×24 = 480	1348	59.9
							25×33 = 825	1693	75.2
	4	8×34= 272	7×12= 84	6×24= 144	6×24= 144	652	0 (no ECC)	714	31.7
							13×12 = 156	930	41.3
							17×24 = 408	1182	52.5
							22×33 = 726	1500	66.7
	3	7×34= 238	6×12= 72	5×24= 120	5×24= 120	558	0 (no ECC)	620	27.6
							11×12 = 132	812	36.1
							15×24 = 360	1040	46.2
							19×33 = 627	1307	58.1
	2	6×34= 204	5×12= 60	4×24= 96	4×24= 96	464	0 (no ECC)	526	23.4
							10×12 = 120	706	31.4
							12×24 = 288	874	38.8
							16×33 = 528	1114	49.5
	1	3×34= 102	0 (note 1)	0 (note 1)	0 (note 1)	110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2
Simple Waveform	0	3×34= 102	0 (note 1)	0 (note 1)	0 (note 1)	110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2

Note 1: The 0 bit means random noise having a unit variance is used for excitation.
Note 2: The total number of VDR bits includes 8 bits for pitch gain and residual peak amplitude.
Note 3: The total number of bits includes 62 bits for MELPe and mode index, the total number of VDR bits, the ECC bits (if applied) on the VDR bits, and the 60 bits of ECC (if applied) on the MELPe bits and mode index.

Table 8 — Mode 5 Quantization Table for Narrowband VDR with ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components)				Total # of VDR Bits (note 3)	Number of VDR ECC Bits for Four ECC Strength Options (t=0,2,4,6)	Total # of Bits (note 4)	Instantaneous Total Bit Rate (kbps)
		Band 1 0.1–1.5 kHz (34)	Band 2 1.5–2.0 kHz (12)	Band 3 2.0–3.0 kHz (24)	Band 4 3.0–4.0 kHz (24)				
Complex Waveform	5	9×34= 306	8×12= 96	7×24= 168		578	0 (no ECC)	640	28.4
							12×12 = 144	844	37.5
							15×24 = 360	1060	47.1
							20×33 = 660	1360	60.4
	4	8×34= 272	7×12= 84	6×24= 144		508	0 (no ECC)	570	25.3
							10×12 = 120	750	33.3
							14×24 = 336	966	42.9
							17×33 = 561	1191	52.9
	3	7×34= 238	6×12= 72	5×24= 120		438	0 (no ECC)	500	22.2
							9×12 = 108	668	29.7
							12×24 = 288	848	37.7
					Not transmitted (note 2)		15×33 = 495	1055	46.9
	2	6×34= 204	5×12= 60	4×24= 96		368	0 (no ECC)	430	19.1
							8×12 = 96	586	26.0
							10×24 = 240	730	32.4
							13×33 = 429	919	40.8
	1	3×34= 102	0 (note 1)	0 (note 1)		110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2
Simple Waveform	0	3×34= 102	0 (note 1)	0 (note 1)		110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2

Note 1: The 0 bit means random noise having a unit variance is used for excitation.
Note 2: The untransmitted spectral components are replicated by the transmitted spectra in the lower bands.
Note 3: The total number of VDR bits includes 8 bits for pitch gain and residual peak amplitude.
Note 4: The total number of bits includes 62 bits for MELPe and mode index, the total number of VDR bits, the ECC bits (if applied) on the VDR bits, and the 60 bits of ECC (if applied) on the MELPe bits and mode index.

Table 9 — Mode 4 Quantization Table for Narrowband VDR with ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components): Band 1 0.1–1.5 kHz (34)	Band 2 1.5–2.0 kHz (12)	Band 3 2.0–3.0 kHz (24)	Band 4 3.0–4.0 kHz (24)	Total # of VDR Bits (note 3)	Number of VDR ECC Bits for Four ECC Strength Options (t=0,2,4,6)	Total # of Bits (note 4)	Instan-taneous Total Bit Rate (kbps)
Complex Waveform	5	9×34= 306	8×12= 96	Not transmitted (note 2)		410	0 (no ECC)	472	21.0
							9×12 = 108	640	28.4
							11×24 = 264	796	35.4
							14×33 = 462	994	44.2
	4	8×34= 272	7×12= 84			364	0 (no ECC)	426	18.9
							8×12 = 96	582	25.9
							10×24 = 240	726	32.3
							13×33 = 429	915	40.7
	3	7×34= 238	6×12= 72			318	0 (no ECC)	380	16.9
							7×12 = 84	524	23.3
							9×24 = 216	656	29.2
							11×33 = 363	803	35.7
	2	6×34= 204	5×12= 60			272	0 (no ECC)	334	14.8
							6×12 = 72	466	20.7
							7×24 = 168	562	25.0
							10×33 = 330	724	32.2
	1	3×34= 102	0 (note 1)			110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2
Simple Waveform	0	3×34= 102	0 (note 1)			110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2

Note 1: The 0 bit means random noise having a unit variance is used for excitation.
Note 2: The untransmitted spectral components are replicated by the transmitted spectra in the lower bands.
Note 3: The total number of VDR bits includes 8 bits for pitch gain and residual peak amplitude.
Note 4: The total number of bits includes 62 bits for MELPe and mode index, the total number of VDR bits, the ECC bits (if applied) on the VDR bits, and the 60 bits of ECC (if applied) on the MELPe bits and mode index.

Table 10 — Mode 3 Quantization Table for Narrowband VDR with ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components)				Total # of VDR Bits (note 2)	Number of VDR ECC Bits for Four ECC Strength Options (t=0,2,4,6)	Total # of Bits (note 3)	Instantaneous Total Bit Rate (kbps)
		Band 1 0.1–1.5KHz (34)	Band 2 1.5–2.0kHz (12)	Band 3 2.0–3.0kHz (24)	Band 4 3.0–4.0kHz (24)				
Complex Waveform	5	9×34= 306	Not transmitted (note 1)			314	0 (no ECC)	376	16.7
							7×12 = 84	520	23.1
							9×24 = 216	652	29.0
							11×33 = 363	799	35.5
	4	8×34= 272				280	0 (no ECC)	342	15.2
							6×12 = 72	474	21.1
							8×24 = 192	594	26.4
							10×33 = 330	732	32.5
	3	7×34= 238				246	0 (no ECC)	308	13.7
							5×12 = 60	428	19.0
							7×24 = 168	536	23.8
							9×33 = 297	665	29.6
	2	6×34= 204				212	0 (no ECC)	274	12.2
							5×12 = 60	394	17.5
							6×24 = 144	478	21.2
							8×33 = 264	598	26.6
	1	3×34= 102				110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2
Simple Waveform	0	3×34= 102				110	0 (no ECC)	172	7.6
							3×12 = 36	268	11.9
							3×24 = 72	304	13.5
							4×33 = 132	364	16.2

Note 1: The 1.5–4.0 kHz is derived not from spectral replication but from that region of the 2.4 kbps MELPe signal.
Note 2: The total number of VDR bits includes 8 bits for pitch gain and residual peak amplitude.
Note 3: The total number of bits includes 62 bits for MELPe and mode index, the total number of VDR bits, the ECC bits (if applied) on the VDR bits, and the 60 bits of ECC (if applied) on the MELPe bits and mode index.

Table 11 — Mode 2 Quantization Table for Narrowband VDR with ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components): Band 1 0.1–0.7 kHz (14)	Band 2 0.7–2.0 kHz (32)	Band 3 2.0–3.0 kHz (24)	Band 4 3.0–4.0 kHz (24)	Total # of VDR Bits (note 2)	Number of VDR ECC Bits for Four ECC Strength Options (t=0,2,4,6)	Total # of Bits (note 3)	Instan-taneous Total Bit Rate (kbps)
Complex Waveform	5	9×14= 126	Not transmitted (note 1)			134	0 (no ECC)	196	8.7
							3×12 = 36	292	13.0
							4×24 = 96	352	15.6
							5×33 = 165	421	18.7
	4	8×14= 112				120	0 (no ECC)	182	8.1
							3×12 = 36	278	12.4
							4×24 = 96	338	15.0
							4×33 = 132	374	16.6
	3	7×14= 98				106	0 (no ECC)	168	7.5
							3×12 = 36	264	11.7
							3×24 = 72	300	13.3
							4×33 = 132	360	16.0
	2	6×14= 84				92	0 (no ECC)	154	6.8
							2×12 = 24	238	10.6
							3×24 = 72	286	12.7
							4×33 = 132	346	15.4
	1	3×14= 42				50	0 (no ECC)	112	5.0
							1×12 = 12	184	8.2
							2×24 = 48	220	9.8
							2×33 = 66	238	10.6
Simple Waveform	0	3×14= 42				50	0 (no ECC)	112	5.0
							1×12 = 12	184	8.2
							2×24 = 48	220	9.8
							2×33 = 66	238	10.6

Note 1: The 0.7–4.0 kHz is derived not from spectral replication but from that region of the 2.4 kbps MELPe signal.
Note 2: The total number of VDR bits includes 8 bits for pitch gain and residual peak amplitude.
Note 3: The total number of bits includes 62 bits for MELPe and mode index, the total number of VDR bits, the ECC bits (if applied) on the VDR bits, and the 60 bits of ECC (if applied) on the MELPe bits and mode index.

3.3 Mode Switching

The capability to dynamically switch between all the VDR modes allows for the efficient use of the communications channel under various and changing conditions. Since these conditions are application specific, the VDR algorithm does not make its own decision on what rate it should be running. It requires external algorithms, perhaps with inputs or feedback from the receiver, to set the rate. This section provides guidelines to some of the issues regarding mode switching, even though most of these issues are separate from the VDR algorithm. Within the VDR algorithm itself, the transmitter just needs to send the current voice frame's mode index for it to be decoded by the receiver.

3.3.1 Rules for Switching Between Modes

One of the main advantages of using a single voice processing principle for the various VDR vocoding modes is that switching between modes is straightforward for most situations. The following is a list of guidelines and capabilities of the VDR algorithm.

- Switching the bit precision *within* a VDR mode based on speech content can be done every frame (as often as 44.44 times per second).

- Switching the error control levels *within* a VDR mode can be done every frame (as often as 44.44 times per second).

- Switching *between* different VDR modes is more complicated because of the presence of lower/upper band filters when using the MELPe upperband in modes 2 and 3 instead of spectral replication found in modes 4, 5, and 6. Switching between VDR modes 4, 5, and 6 can be accomplished every frame. However, switching between mode 2 and any other modes, or between mode 3 and any other modes, should not be done every frame because of the MELPe synthesis filters required. Here it would be more important to switch during periods of silence to avoid speech discontinuities. Downgrading to and upgrading from MELPe mode 1 should not be done every frame, either.

The following three sections present examples of optimizing communication by switching VDR mode based on channel conditions, network congestion, and acoustic noise environment.

3.3.2 Changing VDR Mode Based on Channel Conditions

The many modes listed in Tables 7 through 11 give a large degree of flexibility in adapting to ever-changing channel conditions. But for the transmitter to take advantage of this flexibility, it needs prior knowledge of the channel or feedback from receiver and network nodes. If the transmitter could get feedback from the receiver, it would be best to know when too many errors are not being corrected. Then the transmitter could automatically increase the allocation of error control bits. If necessary, it would even be appropriate in severe channel environments to only use MELPe coding and use the rest of the available bandwidth on error control. In addition, because the goal is to facilitate the possibility of transmitting across multiple dissimilar channel links, each intermediate network link can be formatted with the most appropriate ECC mode.

Therefore, even intermediate nodes could give feedback based on local decoding and correcting of the encrypted bitstream. This is possible because the error control bits are computed based on the already encrypted bitstream. Each network bridge could do this automatically without having to get information all the way back from the final destination. By using received bitstream statistics, it is possible that two-way conversations could give feedback on the overall channel quality each time they transmit and give the preferred mode to the transmitter based on current channel conditions.

If feedback is available, a set of rules for when to increase or decrease ECC is needed. One important consideration is how fast to switch modes based on bitstream errors, taking into account the following:

- The number of uncorrected bit-errors that must accumulate over some number of frames at the receiver before changing to a VDR mode with a higher level of error protection at the transmitter.

- The number of uncorrected bit-errors that must occur at the receiver before changing to a fixed rate MELPe mode with the highest level of error protection (Section 6).

- As fewer bit-errors are encountered at the receiver, the proper threshold to switch back to a mode with less error protection at the transmitter.

Also, it is not necessary for participants in a two-way conversation to transmit in the same mode in each direction. Consider the example of different bilateral modes as when a soldier on shore is communicating with a ship or aircraft. The soldier's radio is less powerful than that of either the ship or the aircraft and thus may be more error prone. In this scenario, although all three may be communicating together, it probably makes sense for the soldier to use a more error-protected, or lower rate (lower power) mode, than either the ship or the aircraft.

3.3.3 Changing VDR Mode Based on Network Congestion

An ideal use for the VDR algorithm is the situation in which a fixed amount of channel capacity is allocated to many users, and each of these users can *variably* go above and below their typical bandwidth allocation. In this situation the VDR algorithm compensates for the fact that allocated spectrum is a perishable resource that cannot be conserved. The variability of the VDR algorithm allows it to produce higher quality speech for a much lower overall average data rate than a fixed rate vocoder, and dynamically varying vocoding efficiently uses the entire allocated spectrum.

In a similar situation involving multiple users sharing a fixed amount of bandwidth, it is possible to maximize the number of users able to access a communications link during an emergency by dynamically lowering individual vocoding rates. As more users want to use the combined channel, individual vocoding rates are reduced so that the combined channel can still support the overall rate. Implicit in this capability is the need for feedback from the system to the transmitters so that VDR modes can be dynamically adjusted based on the number of users at any one time.

3.3.4 Changing VDR Mode Based on Acoustic Noise Environment

Because the DoD needs to operate in severe acoustic noise environments, noise cancellation is already accomplished in the following two ways. First, noise-canceling microphones are standard on all tactical handsets, and second, a noise-canceling preprocessor is part of the NATO standard for MELPe. Since VDR uses MELPe, all modes of VDR have this NC preprocessor built in.

Even with existing noise cancellation, improvements can still be obtained with the VDR algorithm by using more spectral coefficients in harsh acoustic environments. This was verified using speech intelligibility testing (Diagnostic Rhyme Test, DRT) and speech acceptability testing (Diagnostic Acceptability Measure, DAM). This testing has shown that it is certainly advantageous to adapt to severe acoustic environments with increased vocoding levels.

Also, this adaptation could be designed to be platform specific. Some platforms, the CH-46 helicopter for example, may benefit from the setting of a minimum VDR mode. The use of different bilateral modes could also be considered based on the location of each individual speaker in a conversation. Different VDR modes may be necessary because of dissimilar acoustic environments. In the case of someone in a helicopter talking to someone in a combat information center (CIC) of a ship, for example, the acoustic background noise when transmitting from a CH-46 is much louder than the noise experienced when transmitting from the CIC in the ship. So while the MELPe 2400 bps minimum rate may be sufficient when transmitting from the ship, a higher VDR vocoding rate may be needed to overcome the noise when transmitting from the helicopter.

In balancing the combined need to protect against bit errors and against acoustic noise, it is generally better to allocate more ECC bits than vocoder bits. This is because communications systems using NC microphones and the NC preprocessing algorithm already have some protection from acoustic noise, so bit errors are more detrimental to speech quality.

4 EXTENDING VDR TO 16000 BPS FIXED RATE OPTIONS (WITH AND WITHOUT ECC)

One goal of this work is to provide a truly universal vocoder. Ideally, this would be a purely variable rate vocoder, but the predominance of fixed rate channels requires that the variable rate vocoder also contain, or be interoperable with, certain fixed rate vocoding modes.

Important examples of these fixed rate modes are the newly defined modes for tactical secure voice put forth in the Tactical Secure Voice Cryptographic Interoperability Specification (TSVCIS) [4]. TSVCIS is a specification written by the Tactical Secure Voice Working Group (TSVWG) for enabling all modernized tactical secure voice devices to be interoperable across the Department of Defense. A more complete description of TSVCIS is given in Ref. 5. One of the most important aspects is that all the voice modes defined in the TSVCIS are based on a fixed rate variant of NRL's VDR vocoder which uses the MELPe standard as its base. This section describes the fixed rate options at the 16000 bps rate.

In line with the goal of interoperability, two 16000 bps fixed rate modes were designed that use subsets of the complete VDR encoding shown in Table 2. For the fixed rate option, the vocoder uses a fixed, 8-bit residual encoding precision that does not change based on speech complexity or network congestion as it would for the variable encoding option. In all cases, MELPe occupies the first 54 bits of the frame, making for direct secure interoperability between all fixed rate modes.

Because these fixed rate modes are derived from the VDR vocoder, they can also be made interoperable with the variable rate modes by decoding and re-encoding the VDR spectral coefficients. Though the first 54 MELPe bits are interoperable in the black (encrypted bitstream) across all fixed and variable modes, the spectral coefficients for these modes are not able to be transcoded in the black. This is because changing from the 8-bit precision of the fixed rate spectral coefficients to any other precision means de-encrypting the bitstream before re-encoding. Section 5 discusses performance issues of transcoding between variable and fixed rate modes.

4.1 Description of 16000 bps Fixed Rate Modes

To accommodate different channel conditions, two fixed rate 16000 bps voice modes were designed. One mode has embedded error control coding and is intended for use over noisy channels. The other mode has no ECC. It is intended for use over noise-free channels or channels with external ECC.

When using the complete VDR encoding table without variable encoding, the data rate is 36 kbps at the highest precision setting. To adapt the VDR algorithm to encode at a fixed rate of 16000 bps, only subsets of the complete VDR encoding table can be used; this technique was introduced in Section 2.4.3. The fixed rate modes are derived directly from the VDR modes, the only difference being that the bit precision for the spectral coefficients is fixed rather than variable. Table 5 shows a variable mode where only 34 of the 94 coefficients (100 to 1500 Hz band) are sent and with the upper band replaced by the MELPe upper band at the receiver. Table 6 shows a variable mode where only 14 of the 94 coefficients (100 to 700 Hz band) are sent, with the upper band replaced by the MELPe upper band.

Each of these tables is used to derive the two fixed rate modes by fixing the coefficient resolution at 8 bits, no matter the input speech complexity. Table 5 with 34 coefficients set permanently at 8-bit encoding becomes the 16000 bps algorithm *without* ECC shown in Table 12. Table 6 with 14 coefficients

set permanently at 8-bit encoding becomes the 16000 bps algorithm *with* ECC shown in Table 13. By using only 14 coefficients, the voice-coding rate is only approximately 8 kbps, allowing for 8 kbps of ECC to be added to protect the channel from bit errors. In this mode, the MELPe portion of the bitstream is protected by 11 blocks of a BCH (n=15,k=5,t=3) code while the VDR portion is protected by 11 blocks of a Hamming (n=16,k=11,t=1) code with double error detection. Because the MELPe portion of the bitstream is protected much more strongly from bit errors than the VDR portion, better voice quality may sometimes be obtained in severe channel environments by discarding the corrupted VDR spectral coefficients similarly to that described in Section 3.1.

The highlighted areas of the tables show the sections used for the fixed rate modes. Figure 7 shows the spectrum of the spectral coefficients sent for these modes.

Table 12 — VDR Quantization Table for Interfacing to 16000 bps Fixed Rate without ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 1)	Instantaneous Data Rate (kbps)
		0.1–1.5 kHz (34)	1.5–2.0 kHz (12)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×34=306	Not transmitted MELPe used above 1.5 kHz (note 2)			376	
	4	8×34=272				342	15 (note 3)
	3	7×34=238				308	
	2	6×34=204				274	
Fully Unvoiced	1	3×34=102				172	
	0	3×34=102				172	

Note 1: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.
Note 2: The band from 1.5 to 4 kHz is then derived from that region of the 2.4 kbps MELPe signal.
Note 3: The highlighted text shows the portions of Table 5 that are used in the 16000 bps fixed rate without ECC allocation (i.e., all 8-bit encoding)

Table 13 — VDR Quantization Table for Interfacing to 16000 bps Fixed Rate with ECC

Number of Voiced Frequency Bands		Frequency Band in kHz (# of Spectral Components)				Total # of Bits (note 1)	Instantaneous Data Rate (kbps)
		0.1–0.7 kHz (14)	0.7–2.0 kHz (32)	2–3 kHz (24)	3–4 kHz (24)		
Fully Voiced	5	9×14=126	Not transmitted MELPe used above 0.7 kHz (note 2)			196	
	4	8×14=112				182	8 (note 3)
	3	7×14=98				168	
	2	6×14=84				154	
Fully Unvoiced	1	3×14=42				112	
	0	3×14=42				112	

Note 1: The total number of bits includes 70 bits for the MELPe standard, pitch gain, residual peak amplitude, and the operating mode selector.
Note 2: The band from 0.7 to 4 kHz is then derived from that region of the 2.4 kbps MELPe signal.
Note 3: The highlighted text shows the portions of Table 6 that are used in the 16000 bps fixed rate with ECC allocation (i.e., all 8-bit encoding)

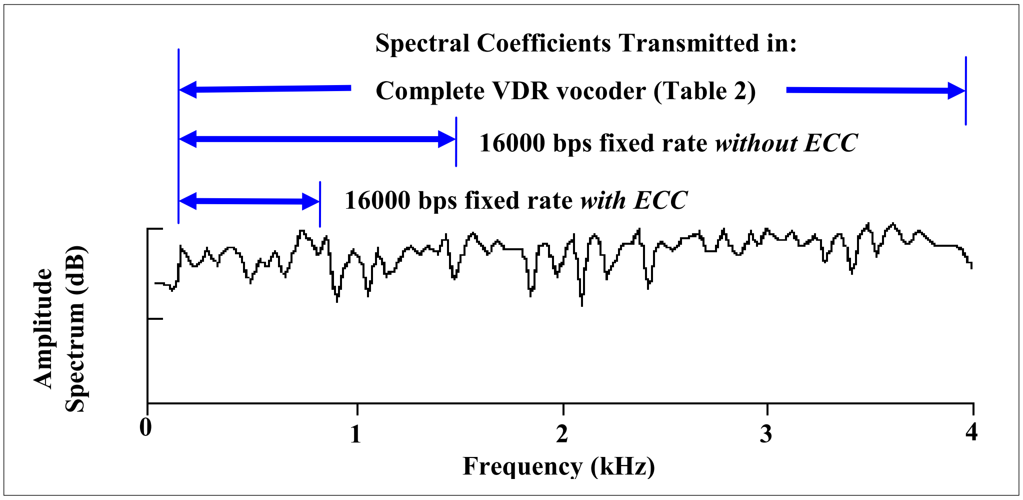


Fig. 7 — Example of the 4 kHz, 96-point residual spectrum and the portion used for a given operating mode, as indicated. The complete VDR system encodes the 100–4000 Hz band. The 16000 bps fixed rate mode without ECC encodes the 100–1500 Hz band. The 16000 bps fixed rate mode with ECC encodes the 100–700 Hz band.

5 TRANSCODING 16000 BPS FIXED RATE MODES TO/FROM VARIABLE RATE MODES

5.1 VDR Modes and Fixed Rate Modes to be Transcoded

To have the fixed and variable rate vocoding modes interoperate, a method must exist for converting the transmitted speech parameters from one mode to the other, in mid-transmission, without reprocessing the original speech samples. Because these versions of the vocoder were all developed from the same voice processing principle, transcoding between modes can be accomplished in a straightforward way. This section describes that method in detail. Six transcoding options are listed below. In addition to these, all modes can be converted to 2400 bps MELPe without any degradation compared with the original MELPe bitstream. This is because all modes include that narrowband bitstream as the first 54 bits of each frame.

Table 14 — Transcoding Options Between VDR and Fixed Rate Modes

Variable Data Rate to Fixed Rate Transcoding Options		
Mode 6 full rate VDR (Table 2)	transcoded to	16000 bps fixed rate without ECC
Mode 3 limited VDR (Table 5)	transcoded to	16000 bps fixed rate without ECC
Mode 6 full rate VDR (Table 2)	transcoded to	16000 bps fixed rate with ECC
Mode 2 limited VDR (Table 6)	transcoded to	16000 bps fixed rate with ECC
Fixed Rate to Variable Data Rate Transcoding Options		
16000 bps fixed rate without ECC	transcoded to	Mode 3 limited VDR (Table 5)
16000 bps fixed rate with ECC	transcoded to	Mode 2 limited VDR (Table 6)

5.1.1 VDR to Fixed Transcoding

The limited VDR mode is a subset of the full VDR mode. The limited rate VDR mode is transcoded to either of the 16000 bps fixed rate modes by converting all that mode's spectral coefficients to fixed, 8-bit values. The full VDR mode is first converted to the limited VDR mode by discarding the upper spectral coefficients and then transcoded to the fixed rates in the same way as the limited VDR mode.

5.1.2 Fixed Rate to VDR Transcoding

The two 16000 bps fixed rate modes do not contain the upper spectral coefficients necessary for the complete VDR spectrum, so they can only be converted to the limited VDR mode. The fixed rate modes cannot be converted to the full VDR mode.

5.2 Conversion Between Different Precision of Spectral Constellations

All the transcoding between modes listed above is based on converting the prediction residual spectral coefficients between the different spectral constellations. The constellations in Fig. 4(a) (9-bit, 512-point) and Fig. 4(b) (7-bit, 128-point) are two examples that show how these different precisions are mapped onto the spectral plane.

When converting VDR to a fixed rate mode, the varying bit precision of the spectral coefficients get mapped to a fixed, 8-bit precision. This change in precision is a decrease for fully voiced frames and an increase for unvoiced frames. This suggests that some voiced frames may suffer degradation in quality from the conversion but unvoiced frames should not.

When converting a fixed rate mode to a VDR mode, the degree of voicing is used to determine the bit precision of the VDR mode spectral coefficients. It is the opposite of the process above.

The continuous task of converting an incoming residual point from one table to another by going through each point in each table and finding its "nearest neighbor" is CPU intensive. Fortunately these calculations can all be precalculated and saved to memory as a "mapping" table. This conversion is a simple process because all these modes are designed with the same voice processing parameters, just with varying precision.

Although the amount of degradation caused by the conversion process can be measured by calculating the spectral difference from one spectral point to another, it is more important to measure the perceived error. The human ear is less sensitive to errors at higher frequencies, so testing methodologies are used that take this into account. The VDR algorithm takes advantage of this fact by encoding higher frequencies with lower resolution.

5.3 Testing Sample

When evaluating the performance of this transcoding process, it is important to have a representative speech sample. The voice sample we used is an approximately 37-second conversation between a male and a female speaker, with typically small gaps between speakers' questions and answers. Figure 8 shows the distribution of the degree of voicing in the 1656 frames analyzed. (Each frame is 22.5 ms of speech.) The degree of voicing is important because this is how the bit precision of the spectral coefficients is determined. By analyzing each frame of speech, it was found that in this 37-second speech recording, VDR changed the data rate 818 times based on the degree of voicing. This tremendous data rate flexibility is what makes VDR vocoding a very good technique for adapting to changing conditions.

This graph also shows the wide distribution of the various bit allocations. While 9-bit encoding is optimum for speech quality, it was only needed in approximately 20% of the speech frames. In fact, the average bit precision of this entire voice sample encoded with VDR is approximately 5.4 bits, while the performance is at least as good as the fixed rate systems that allocate 8 bits to all spectral coefficients.

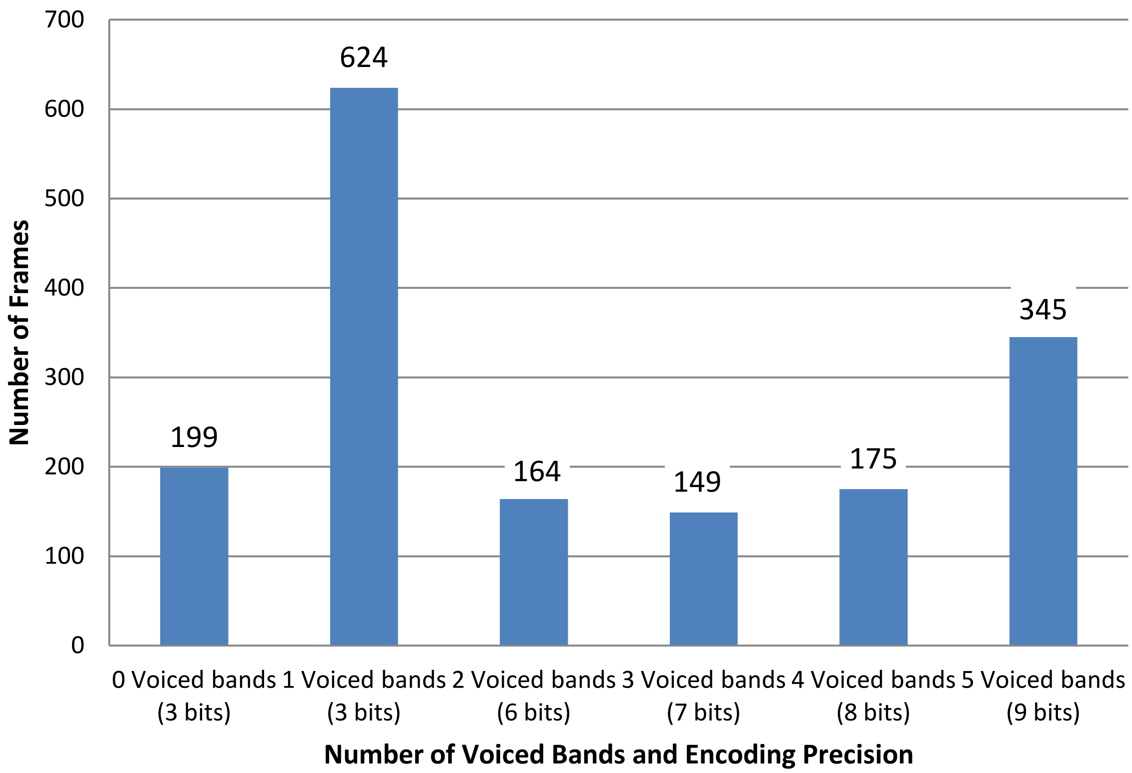


Fig. 8 — The distribution of VDR's spectral coefficient bit allocation for 1656 speech frames (in a 37-second conversation)

5.4 Transcoding from Variable Data Rate Modes to Fixed Rate Modes

Sections 5.4.1 through 5.4.4 describe the process of transcoding from VDR modes to the fixed rate modes. In each full rate VDR mode conversion, the mode is first downgraded to the limited VDR case by discarding the upper spectral coefficients. So as mentioned above, the two limited rate transcoding options are subsets of the full rate conversions. The main conversion process in each of these cases entails converting all the spectral coefficients to 8-bit precision. While there is some precision lost going from 9 bits to 8 bits (for fully voiced speech), 8 bits was seen as a minimum precision level for encoding voiced speech at a good quality. In upconverting the partially unvoiced and silent speech frames to 8 bits, there should be very little degradation because of the closeness of the 256 points available in the constellation.

5.4.1 Full Rate VDR (Table 2) Transcoded to 16000 bps Fixed Rate without ECC

As shown in Fig. 9, the full rate VDR mode encodes 94 coefficients at 3-bit to 9-bit precision based on the degree of voicing. To convert this to the fixed mode without ECC, all the coefficients above coefficient 34 are first discarded. This effectively converts this to the limited VDR case. Then, the 34 coefficients are re-encoded at 8-bit precision. Finally, the MELPe upperband is inserted as normally done.

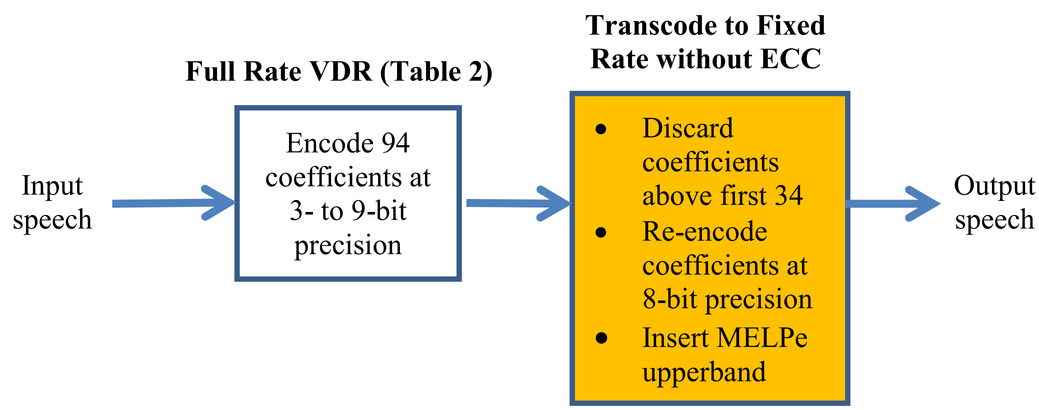


Fig. 9 — The process of transcoding from full rate VDR (Table 2) to fixed rate without ECC mode. The yellow block is the transcoder.

5.4.2 Limited VDR (Table 5) Transcoded to 16000 bps Fixed Rate without ECC

As shown in Fig. 10, the limited VDR (Table 5) mode encodes 34 spectral coefficients at 3- to 9-bit precision based on the degree of voicing. This conversion is just a subset of the process shown in Fig. 9 where the spectral coefficients are then re-encoded at 8-bit precision, with the MELPe upperband inserted as normal.

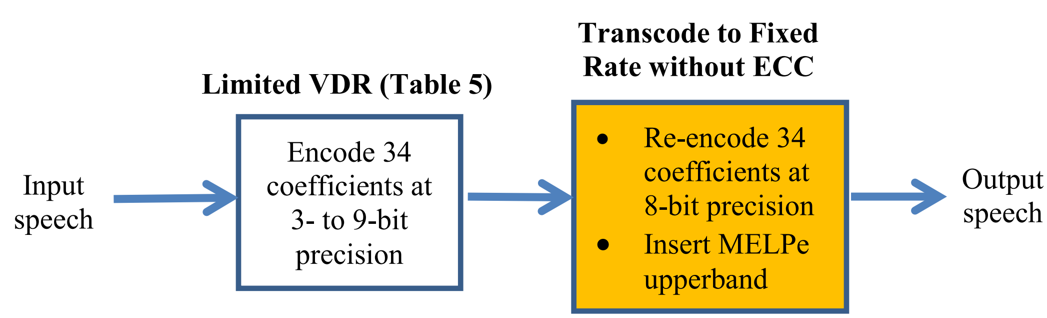


Fig. 10 — The process of transcoding from limited VDR (Table 5) to fixed rate without ECC mode. The yellow block is the transcoder.

5.4.3 Full Rate VDR (Table 2) Transcoded to16000 bps Fixed Rate with ECC

As shown in Fig. 11, this transcoding process is identical to the conversion in Section 5.4.1, except there are only 14 coefficients to be converted. This reduction in the number of spectral coefficients from 34 to 14 frees space for the error control bits in this mode. After discarding the upper coefficients and then converting the remaining coefficients to 8-bit precision, the MELPe upperband is inserted as normal.

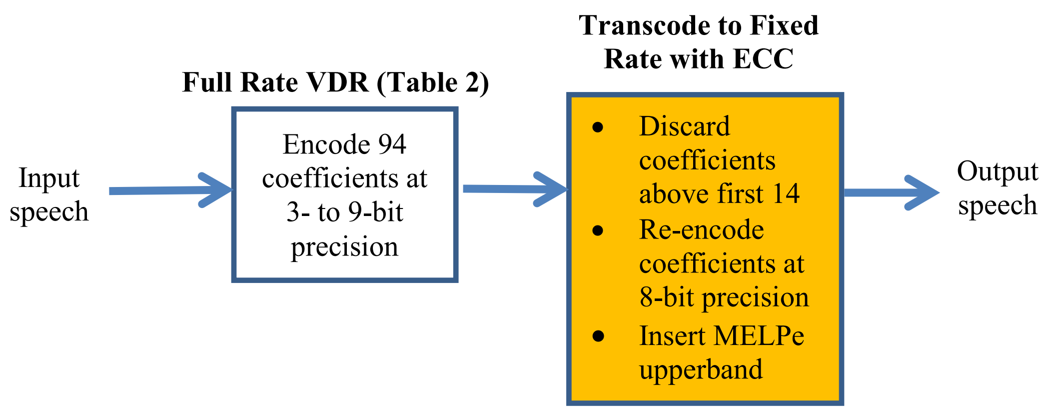


Fig. 11 — The process of transcoding from full rate VDR (Table 2) to fixed rate with ECC mode. The yellow block is the transcoder.

5.4.4 Limited VDR (Table 6) Transcoded to 16000 bps Fixed Rate with ECC

As shown in Fig. 12, this transcoding process is identical to the conversion in Section 5.4.2, except with 14 spectral coefficients instead of 34. Again, the reduction in the number of spectral coefficients frees space for the error control bits within the frame. Also, this is a subset of the full rate VDR conversion above.

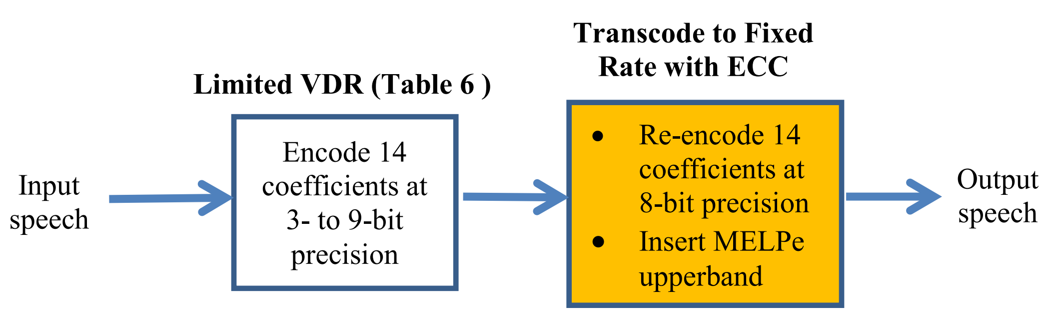


Fig. 12 — The process of transcoding from limited VDR (Table 6) to fixed rate with ECC mode. The yellow block is the transcoder.

5.4.5 Results for Transcoding from VDR to Fixed Rate Modes

The only difference in the vocoding between the VDR modes and fixed rate modes is in the encoding of the spectral coefficients. In the VDR modes, these are encoded with a precision that ranges from 3 to 9 bits. The precision varies frame by frame and depends on the voicing decision. When transcoding to the fixed rate modes, these varying coefficients get re-encoded to a fixed precision of 8 bits. When evaluating the performance of those two conversions, it is the degradation caused by this now two-stage encoding of the spectral coefficients that needs to be measured.

To judge the quality of the transcoded output speech, there are two main cases to be measured in converting to 8-bit precision for all spectral coefficients. In one case the original encoded frame is fully voiced and encoded at 9-bit precision. The other case is when the speech in the original frames is predominantly or fully unvoiced and encoded at 3-bit precision. (In fact, there is a third case of converting mixed voiced frames, but the encoding constellations between 6-, 7-, and 8-bit constellations are close enough to each other to keep conversion errors to a minimum.)

With a fully voiced frame, the spectral coefficients of the original speech that are encoded with a precision of 9 bits are transcoded to a precision of 8 bits. This transcoding results in a very slight degradation of the original speech. This is due to the small distance the quantized spectral coefficients are shifted when moving them from the 512-point constellation used for 9-bit coefficients to the 256-point constellation used for 8-bit coefficients.

When comparing the transcoded speech to speech that was originally encoded with 8-bit spectral coefficients, there is essentially no degradation. In other words, fully voiced speech that is encoded with the fixed rate mode has the same voice quality as speech that was first encoded with the VDR mode and then transcoded to that same fixed rate mode.

In the fully or predominantly unvoiced case, a consonant originally encoded at 3 bits gets re-encoded at 8 bits. This increase in bits does not equate to an increase in precision, as the information was lost in the original 3-bit encoding. So there is no effect on voice quality when transcoding fully or predominantly unvoiced frames.

When speech is originally encoded in the fixed rate mode, fully unvoiced frames are directly encoded with 8 bits. This increase in precision, compared to the VDR modes, provides very little, if any, improvement in performance in the consonants. In other words, the decision by the VDR systems to reduce the precision of spectral coefficients for consonants because they are much like random noise was a good one.

5.5 Transcoding from Fixed Rate Modes to Variable Data Rate Modes

Sections 5.5.1 and 5.5.2 describe the process of transcoding from the fixed rate modes to the VDR modes. Because the MELPe voicing data is available to the receiver, the transcoder can convert these fixed rate modes back to variable modes, if so desired. This entails converting the 8-bit constellations back to the 3- to 9-bit constellations called for by the degree of voicing in the speech.

From the standpoint of interoperability, however, these conversions would not be required if the receiving network had these fixed rate modes built into its system. The receiving network could simply transmit and synthesize the 16000 bps fixed rate voice signal. But if a traditional high bandwidth network wants to have only VDR modes, these conversions are necessary so the network can take advantage of significantly reducing the data rate when there are many speech gaps in an “always on” open circuit conversation.

5.5.1 16000 bps Fixed Rate without ECC Transcoded to Limited VDR (Table 5)

As shown in Fig. 13, the 16000 bps fixed rate without ECC vocoder first encodes all 34 coefficients at 8-bit precision. The transcoding process then converts all 34 coefficients to the 3- to 9-bit precision based on the degree of voicing present in the frame.

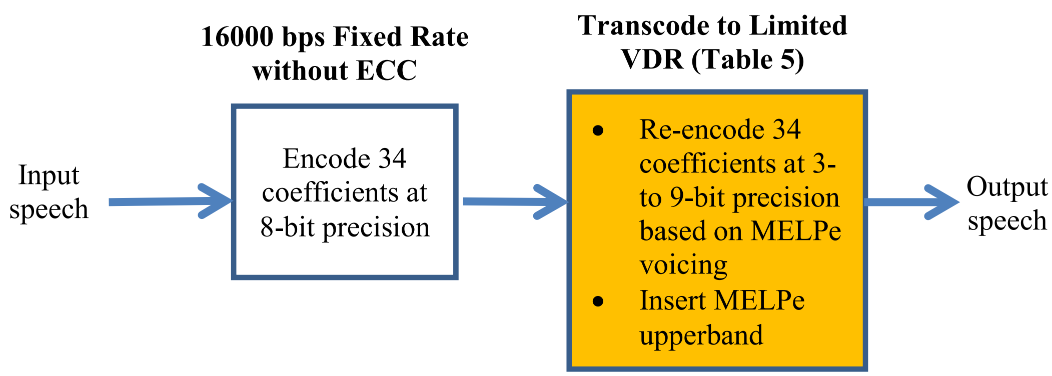


Fig. 13 — The process of transcoding from the fixed rate without ECC mode to the limited VDR (Table 5) mode. The yellow block is the transcoder.

5.5.2 16000 bps Fixed Rate with ECC Transcoded to Limited VDR (Table 6)

As shown in Fig. 14, this case is identical to that in Section 5.5.1 except with 14 coefficients instead of 34 coefficients.

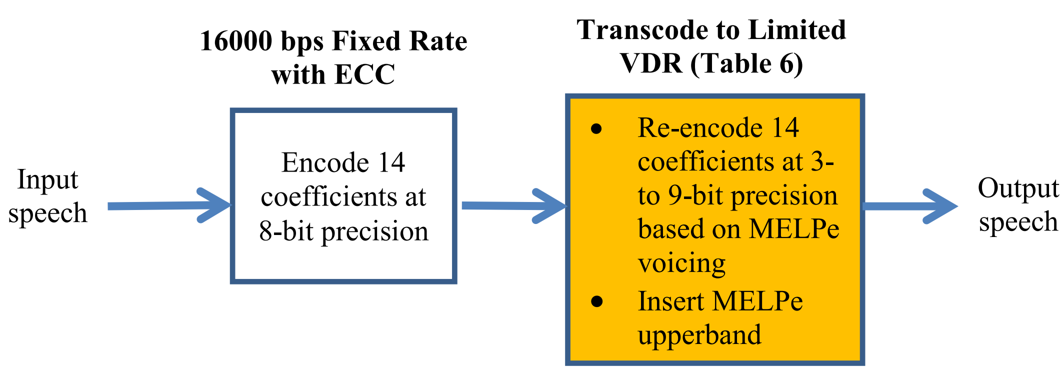


Fig. 14 – The process of transcoding from the fixed rate with ECC mode to the limited VDR (Table 6) mode. The yellow block is the transcoder.

5.5.3 Results for Transcoding from Fixed Rate to VDR Modes

As with the results presented in Section 5.4.5 (transcoding VDR to fixed rate), the judgment to make here is how much voice quality is lost by the two-stage encoding process compared with sending the fixed rate modes all the way through the system. Also, it is again best to focus on the two main cases of converting to fully or predominantly unvoiced (3-bit spectral encoding) or completely voiced speech (9-bit spectral encoding).

For the unvoiced case, when downgrading the 8-bit constellation to a lower precision, it again becomes apparent that there is very little performance degradation in the consonants because 8-bit encoding for unvoiced speech is not necessary for good sounding consonants. For the fully voiced case, in converting the 8-bit constellation up to 9 bits, again there is no degradation in speech quality. Keep in mind that the speech quality is not enhanced either, though, for the same reason that upconverting a traditional DVD to high definition (HD) television does not make it HD quality. So while it may be advantageous to convert back to the limited VDR modes for bandwidth concerns (especially in channels where there are commonly many silent gaps), voice quality will not be improved by doing so.

6 DESIGNING FIXED RATE 8000, 12000, 600, AND 1200 MELPE MODES WITH ECC INTO BIT ERROR TOLERANT MODES

This section describes four additional modes that add to the universal nature of the vocoder. They are fixed rate options directly based on the MELPe vocoding standard alone without any additional spectral coefficients added. The rates of these four modes are 8000 bps, 12000 bps, and two options at 2400 bps. These modes were written into the TSVCIS to give system developers fixed rate options in very error prone channels. In essence, these modes are just heavily error protected versions of various MELPe options.

6.1 8000 and 12000 bps Fixed Rate Modes Based on 2400 bps MELPe Option

The 8000 bps mode (180 bits per 22.5 ms speech frame) appends ECC bits onto the 54-bit MELPe 2400 bps bitstream. The ECC method used is BCH (n=15,k=5,t=3) so that the MELPe bitstream is encoded by appending 10 ECC bits to every 5 MELPe bits. Because there are 54 bits in the MELPe bitstream, with a 55th bit reserved for future use, this process is repeated 11 times to form 165 bits. Other control bits are added to form the 180-bit bitstream which translates to 8000 bps. This process is shown in Fig. 15.

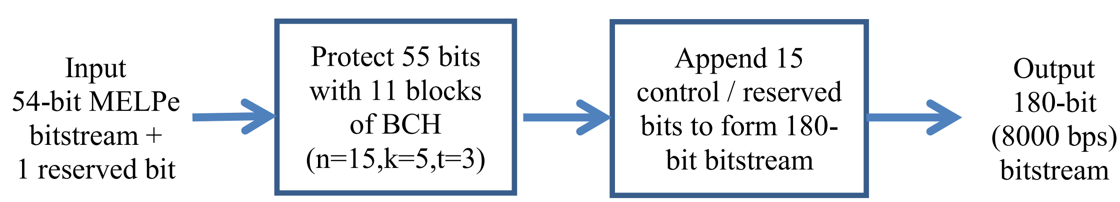


Fig. 15 — Frame-by-frame conversion to 8000 bps fixed rate mode from the MELPe bitstream

To form the 12000 bps mode (270 bits per frame), another layer of ECC coding is added onto the 55-bit bitstream in addition to the 8000 bps mode. This coding is BCH (n=125,k=55,t=11) and the idea of this mode is to help correct the remaining errors in the MELPe bitstream after the BCH (n=15,k=5,t=3)

has been decoded. As this encoding adds 70 bits to the bitstream, more capacity is available for more control bits in the 270-bit bitstream. This process is shown in Fig. 16.

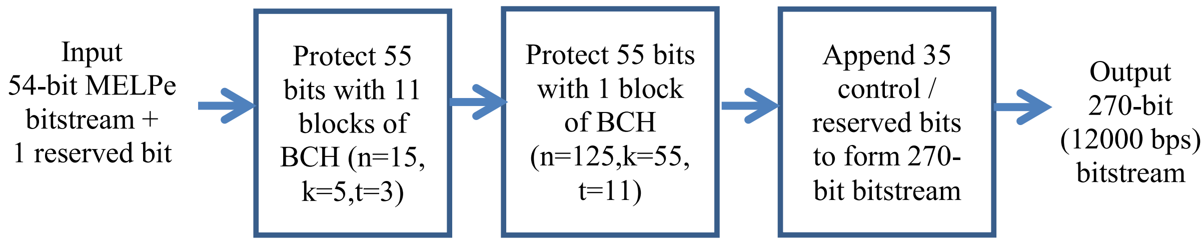


Fig. 16 — Frame-by-frame conversion to 12000 bps fixed rate mode from the MELPe bitstream

The advantage of these two modes is that they are directly interoperable with the 2400 bps MELPe option because the only vocoding bits are the first 54 bits of the 2400 bps MELPe vocoder. In other words, in the presence of no bit errors, the vocoding performance is exactly the same as the 2400 bps MELPe option. But when there are significant amounts of bit errors, performance is improved significantly by this error correction [6].

6.2 2400 bps Fixed Rate Modes Based on 1200 and 600 bps MELPe Vocoding Options

These two modes both have a final rate of 2400 bps but use two different modes of the MELPe standard to get there. One mode uses the 1200 bps MELPe mode and the other uses the 600 bps MELPe mode. Both of these 2400 bps modes involve adding ECC to the bitstream to form a much more bit error tolerant 2400 bps mode. The disadvantage of these two modes is that they are not directly compatible with the 2400 bps MELPe mode because they use a superframe vocoding method to achieve such low vocoding rates. These two modes are the only modes discussed in this report that do not include the first 54 bits of the 2400 bps MELPe bitstream to allow direct compatibility with any communication equipment that uses the 2400 bps MELPe standard.

The first 2400 bps bit error tolerant mode (1200/2400) is based on adding ECC to the 1200 bps MELPe option to form the 2400 bps bitstream. Typically, the voice frames described in this report are 22.5 ms long. To achieve the 1200 bps rate, this MELPe option uses a three-frame superframe so that 81 vocoding bits are transmitted over a 67.5 ms period. These 81 bits are then protected by three blocks of BCH (n=54,k=27,t=5) coding. This process is shown in Fig. 17.

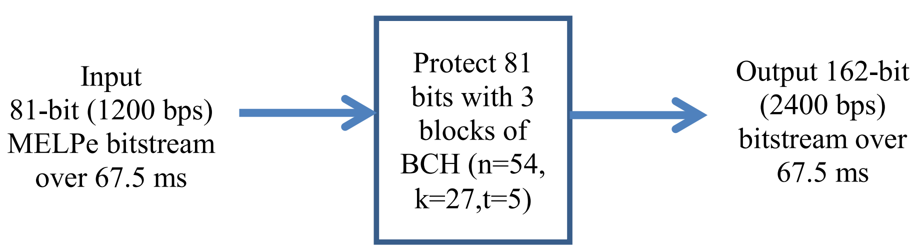


Fig. 17 — Frame-by-frame conversion from 1200 bps MELPe mode to 1200/2400 bps fixed rate mode

The second 2400 bps bit error tolerant mode (600/2400) is based on adding ECC to the 600 bps MELPe option to form the 2400 bps bitstream. To achieve the 600 bps rate, this MELPe option uses a four-frame superframe so that 54 vocoding bits are transmitted over a 90 ms period. One additional sync bit is appended to total 55 bits. These 55 bits are protected by 11 blocks of BCH (n=15,k=5,t=3). A second layer of BCH (n=104, k=55, t=7) is added to correct additional errors not corrected by this first layer of error protection. This process is shown in Fig. 18.

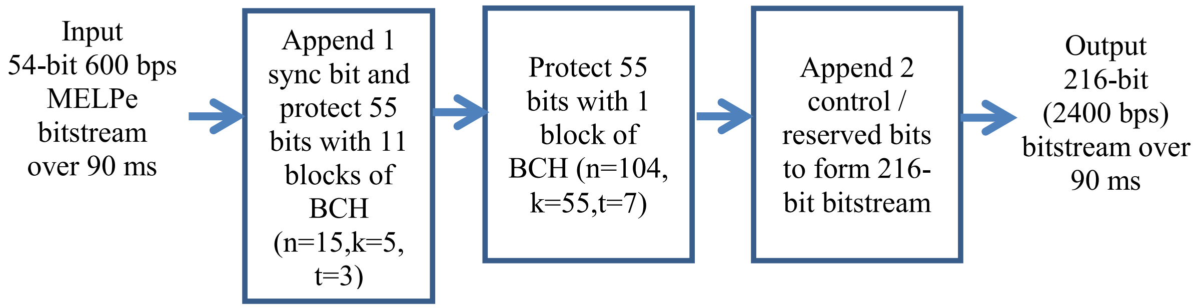


Fig. 18 — Frame-by-frame conversion from 600 bps MELPe mode to 600/2400 bps fixed rate mode

7 CONCLUSIONS

This report documents advancements NRL has made in the effort to achieve a universal vocoder for DoD applications. Four of the most important improvements are these:

- Significant improvements to the VDR vocoder make it much more robust in the less than ideal environments where it may need to operate.
- Error control coding is now extended to all VDR modes; now many more voice applications can be protected in difficult channel environments. This expansion gives system developers many more options to meet their communication requirements.
- Fixed rate vocoding modes based directly on the VDR encoding method were designed. These fixed rate modes are essential for some DoD applications, and by basing them on VDR, transcoding between these options can be done directly and with very little degradation in voice quality.
- Heavily error protected, fixed rate MELPe modes were designed. These modes can be used as fail-safe modes to ensure communicability when channel conditions deteriorate to previously unusable levels.

ACKNOWLEDGMENTS

This work was sponsored by NRL's base research program. The authors wish to thank this program for its strong support of voice processing research over the years. The authors also wish to thank Ha Tang of the U.S. Navy's Program Executive Office C4I and Space, and the National Security Agency's Rick Cantu, chairman of the TSVWG. Finally, the authors would like to thank NRL's Dr. Sastry Kompella for reviewing the report.

REFERENCES

1. T.M. Moran, D.A. Heide, Y.T. Lee, and G.S. Kang, “Variable Data Rate Voice Encoder for Narrowband and Wideband Speech,” NRL/FR/5555--07-10,145, Naval Research Laboratory, Washington, DC, March 2, 2007.

2. G.S. Kang, “Variable-Data-Rate Voice Encoder for Voice Over Internet Protocol (VoIP),” NRL/FR-MM/5550--01-10,016, Naval Research Laboratory, Washington, DC, December 28, 2001.

3. “The 600 Bits/s, 1,200 Bits/s, and 2,400 Bits/s NATO Interoperable Narrow Band Voice Coder,” NATO Standardization Agreement STANAG 4591, NATO Standardization Agency, Brussels, Belgium, January 25, 2006.

4. “Tactical Secure Voice Cryptographic Interoperability Specification (TSVCIS) Version 2.1,” Tactical Secure Voice Working Group (TSVWG), July 2, 2012.

5. T. Moran, D. Heide, and S. Shah, “An Overview of the Tactical Secure Voice Cryptographic Interoperability Specification,” Proceedings of the MILCOM 2010 Military Communications Conference, pp. 213–218 (IEEE, Piscataway, NJ, 2010).

6. P. Shahan, D.A. Heide, and A.E. Cohen, “Comparison of TSVCIS Voice at 8000 and 12000 bps Versus CVSD at 16000 bps,” Proceedings of the MILCOM 2012 Military Communications Conference; doi:10.1109/MILCOM.2012.6415564.